HELLO EVERYONE!

Machine learning project presentation
15.12.2021
PROJECT TEAM AND OWNER

- Margus Lehtme – Msc Data science student
- Tauri Viil – Phd student
- Kwasi A. Boateng – Msc Robotics and computer engineering student
- Rasmus Bobkov - Msc Data science student

- Owner

Sulev Reisberg - Research Fellow of Health Informatics, Chair of data science (Phd)
Microfilmed old church books

Value to be gained

- Estonian national archive has thousands of books like this!

- Valuable historical information about Estonian family- and blood relations

- How many other handwritten books does Estonian national archive have 😊?
I/O WHAT WE HOPE TO ACHIEVE

Input
Identify Surnames?

Output as an index

- SURNAME1, LINK TO FRAME1, TITLE OF THE DOCUMENT1
- SURNAME2, LINK TO FRAME2, TITLE OF THE DOCUMENT2
- SURNAME_N, LINK TO FRAME_X, TITLE OF THE DOCUMENT_Y
MVP

Open source convolutional network architecture

Transkribus®

40 pages of annotated data

VGG image annotator (VIA)
RESULTS

GREAT SUCCESS!

WUT
EVEN MORE RESULTS

Weights & Biases
DOUBLING PIXELS
NOW WITH AUGMENTING THE DATASET
<table>
<thead>
<tr>
<th>Name</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>StAZH_RRB_German_Kurrent_XIX</td>
<td>German</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete training of all minutes of the handwritten Zurich executive minutes (1803-1887), based on automated text-to-image processing.</td>
<td>Nr. of Epochs 400</td>
</tr>
<tr>
<td>The set is licensed under CC-BY-SA and can be re-used.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Document Type</th>
<th>Nr. of Words</th>
<th>Nr. of Lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handwritten</td>
<td>26026908</td>
<td>4617703</td>
</tr>
</tbody>
</table>

**Breaking News!**
ABSOLUTELY NOTHING HAPPENED!
THANK YOU FOR YOUR ATTENTION!

Machine learning project presentation
15.12.2021